
Quality of abstracts of original research articles in *CMAJ* in 1989

Lutchmie Narine, MSc; Deanna S. Yee, BScPhm; Thomas R. Einarson, PhD;
A. Lane Ilersich, BScPhm

Purpose: To evaluate the quality of abstracts of original research articles.

Design: Blind, criterion-based survey.

Sample: Systematic sample of 33 abstracts of original research articles published in *CMAJ* in 1989.

Measurement: The quality of abstracts was measured against a checklist of evaluation criteria, which were divided into eight categories. A score for each abstract was obtained by dividing the number of criteria present by the number applicable. The overall mean score was also determined.

Results: The overall mean score of abstract quality was 0.63 (standard deviation 0.13) out of 1. Of the abstracts reporting study design 56% did not include specific technical descriptors. About 52% did not explicitly describe the study variables. In describing subject selection 79% failed to use specific technical terms. Of the abstracts reporting results 66% did not provide appropriate supporting data. Of those that gave conclusions 86% did not address study limitations and 93% made no recommendations for future study.

Conclusion: Most of the abstracts provided some information pertaining to each evaluation criterion but did not provide detail sufficient to enhance the reader's understanding of the article. On the basis of the study sample the abstracts need improvement in description of research design, reporting of subject selection and results, and statements of limitations and recommendations. The small sample from one journal and the absence of comparison between the contents of the abstracts and the contents of the articles were limitations. Future studies should address these issues and compare the quality of traditional and structured abstracts.

Objet : Évaluer la qualité des résumés d'articles sur des recherches originales.

Conception : Sondage aveugle fondé sur des critères.

Échantillon : Échantillon systématique de 33 résumés d'articles sur des recherches originales publiés dans le *JAMC* en 1989.

Mesure : On a mesuré la qualité des résumés en fonction d'une liste de critères d'évaluation répartis en huit catégories. On a calculé une note pour chaque résumé en divisant le nombre de critères présents dans celui-ci par le nombre de critères applicables. On a aussi calculé le résultat moyen global.

Résultats : Le résultat moyen global de la qualité des résumés a été de 0,63 (écart type de 0,13) sur 1. Parmi les résumés décrivant la conception de l'étude, 56 % ne comportaient pas de descripteurs techniques précis. Environ 52 % ne décrivaient pas explicitement les variables de l'étude. En décrivant la sélection des sujets, 79 % des articles ont omis d'utiliser des termes techniques spécifiques. Parmi les résumés faisant état des résultats, 66 % ne contenaient pas de données d'appui appropriées. Parmi les résumés qui présentaient des conclusions, 86 % ne tenaient pas compte des limites de l'étude et 93 % ne faisaient aucune recommandation relative aux études futures.

Mr. Narine is a PhD candidate in the Department of Health Administration, Faculty of Medicine, University of Toronto, Toronto, Ont. Ms. Yee and Mr. Ilersich are in the master's program in the Faculty of Pharmacy, University of Toronto. Dr. Einarson is an assistant professor in the Department of Health Administration and the Faculty of Pharmacy, University of Toronto.

Reprint requests to: Mr. Lutchmie Narine, Department of Health Administration, Faculty of Medicine, University of Toronto, McMurrich Building, 2nd floor, 12 Queen's Park Cresc. W, Toronto, ON M5S 1A8

Conclusion : La plupart des résumés contenaient quelques renseignements liés à chaque critère d'évaluation, mais ne fournissaient pas suffisamment de détails pour aider le lecteur à comprendre l'article. Si l'on se base sur l'échantillon étudié, il faut améliorer les aspects suivants des résumés : description de la conception de la recherche, rapports sur la sélection des sujets et les résultats, énoncés des limites et recommandations. Le sondage a été limité par l'échantillon restreint de résumés tirés d'une revue et par l'absence de comparaison entre le contenu des résumés et celui des articles. À l'avenir, les études devraient analyser ces questions plus à fond et comparer la qualité des résumés traditionnels et structurés.

Abstracts of published manuscripts were introduced in the 1950s.¹ They were brief summaries of the article and appeared at the beginning. Abstracts were intended to replace the summary that was usually found at the end of the article and to provide the reader with a sense of direction of the article.^{2,3} In the early 1960s *CMAJ* was one of the first journals to adopt the policy of including abstracts.^{4,5} Most major scientific reports in the biomedical literature now begin with a summary or an abstract.

As the volume of biomedical literature increases, readers must become more selective in their choice of reading material. A busy health care professional trying to keep abreast of the latest information will likely peruse the abstract to determine whether to read the entire article. A recent study on MEDLINE use in clinical settings found that abstracts alone often influenced clinical decisions.⁶ An abstract is thus important as a brief but informative synopsis of the article.

Some studies of the quality of traditional abstracts have detected deficiencies. There is some evidence that abstracts tend to present conclusions that do not follow from the findings and that they overemphasize positive conclusions. Gotzsche⁷ investigated overt and hidden bias in 196 reports of double-blind trials of nonsteroidal anti-inflammatory drugs and found that 76% of the reports had doubtful or invalid statements in their conclusions or abstracts. Evans and Pollock⁸ evaluated the shortcomings of 45 articles on antibiotic prophylaxis in surgery. They reported that 20% of the abstracts had either omitted important numerical results or had made unjustified conclusions. In another study Pocock, Hughes and Lee⁹ found bias in the selection of results for inclusion in the abstract of reports of various clinical trials: 70% of the significant findings were included in the abstract, as compared with only 25% of the nonsignificant findings.

We performed this study to evaluate the quality of traditional abstracts in the current medical literature. Our objectives were (a) to develop a valid checklist against which the quality of an abstract could be measured and (b) to evaluate a sample of articles in a medical journal using the instrument developed.

Methods

Sample selection

The journal selected for this study had to contain research articles on health care and require an introductory abstract of those articles. *CMAJ* was selected because it met these criteria and was among the first biomedical journals to adopt the use of abstracts. As well, it was readily accessible to the investigators, and its articles were considered to be representative of biomedical articles.

Only abstracts of original research articles were selected. The reasons for this were that (a) most biomedical articles may be categorized as original research, (b) original research articles are an important source of new knowledge for health professionals and (c) previous work has been done in developing guidelines for the evaluation of abstracts of original research articles.¹⁰

We selected a systematic sample of 33 abstracts of original research articles published in 1989. The subjects addressed in the studies are shown in Table 1. A research assistant photocopied the abstracts. To minimize bias in the evaluation the title, any information pertaining to the authors, the location of the research and the source of any financial support were removed, as suggested by Sacks and associates.¹¹

Evaluation criteria

We then prepared guidelines outlining the infor-

Table 1: Subjects of original research articles published in *CMAJ* in 1989

Subject	No. of articles
Provision or quality of medical services	12
Screening and diagnostic tests	5
Medical intervention	4
Disease causation	4
Disease prognosis	3
Clinical features of a disease	3
Economic analysis	2
Total	33

mation that should appear in an abstract of an original research article. These guidelines were adapted from a number of sources.^{2,10,12,13} The criteria were divided into eight categories, which were labelled as follows: purpose, research design, setting, subjects, intervention, measurement, results and conclusions.

The key details of each section were identified and criteria selected for each important item. Each criterion was represented by a question on a checklist, against which the abstracts were evaluated (Table 2). Each question was weighted equally and could be answered as Yes, No or Not applicable.

Assessment

The abstracts were reviewed independently by two of us (L.N. and D.S.Y.). Communication between the raters was avoided during evaluation. Pilot testing of the checklist was done on two samples of 10 abstracts of original research articles published in *CMAJ* in 1987; the abstracts were randomly selected by the research assistant.

Abstract quality (Q) was defined as the proportion of criteria present according to the formula $Q = Y \div (Y + N)$, where Y was the number of Yes and N the number of No answers; Not applicable responses

Table 2: Criteria used to evaluate the abstracts of the articles

Criterion	Yes	No	Not applicable
Purpose			
Was any information on the purpose given?	31	2	—
Was the purpose explicitly stated?	26	5	2
Was the main purpose distinguished from secondary ones?	—	—	33
Research design			
Was any information on the research design given?	25	8	—
Were technical descriptors used?	11	14	8
If a follow-up study was the duration given?	3	1	29
Setting			
Was any information on the setting given?	18	15	—
Was the level of clinical care (e.g., primary care) indicated?	13	5	15
Subjects			
Was any information on the subjects given?	29	4	—
Were common demographic characteristics given?	27	2	4
Were technical descriptors of subject selection (e.g., random or convenience sample) used?	6	23	4
Was the number of subjects indicated?	25	4	4
Were the response and refusal rates indicated?	2	24	7
Was the number of dropouts and losses indicated?	1	4	28
If the samples were matched were matching characteristics given?	3	1	29
Intervention			
Was any information on intervention given?	4	—	29
Were the commonest name and common synonyms given?	4	—	29
Was a description given?	2	2	29
Was the duration indicated?	3	1	29
Measurement of variables			
Was any information on the measures given?	32	1	—
Were the variables explicitly identified?	15	17	1
Was the source of the data given?	20	12	1
If the measurements were subjective were the observers blind to the patient groupings?	2	—	31
Results			
Were any results given?	32	1	—
Were they directly related to the purpose?	30	—	3
Were appropriate numerical data given?	11	21	1
Conclusions			
Were any conclusions made?	29	4	—
Were they directly related to the purpose?	21	6	6
Were they consistent with the results?	10	19	4
Were the study's limitations mentioned?	4	25	4
Were the study's implications mentioned?	17	12	4
Were there recommendations for further study?	2	27	4

were omitted from the calculation. Each abstract could achieve a maximum quality score of 1 and a minimum score of 0. An overall mean score for the sample was calculated.

The evaluation forms were examined by the research assistant to identify discrepancies between the evaluators. If there was a disagreement the assistant acted as a referee and independently assessed the criterion; the decision was considered to be final.

Interrater reliability

Interrater reliability in the pilot study was determined to ensure that the two evaluators concurred in their application and interpretation of the criteria. It was also determined in the actual study.

The method used to calculate the interrater reliability had been outlined by Rosenthal.¹⁴ It corrects for attenuation in correlations between raters owing to small samples. In this study a Pearson's product moment correlation coefficient was obtained between the scores of the two raters for all the abstracts. The effective interrater reliability was then calculated by substituting the obtained Pearson's r in a modified Spearman-Brown formula.

Results

Instrument reliability

The pilot test of the checklist achieved low agreement because of a systematic difference in interpretation between the raters ($r = 0.42$ with 8 degrees of freedom [df], $p = 0.22$, effective interrater reliability = 0.59). The problems were identified and the questions reworded to be more precise. A second pilot test with the modified instrument produced high agreement ($r = 0.68$ with 8 df, $p = 0.03$, effective interrater reliability = 0.81). Interrater reliability in the actual study was also high, 0.82 ($r = 0.69$ with 31 df, $p < 0.001$); this indicated that there was acceptable agreement among the raters.

Abstract quality

The overall mean score of the abstracts was 0.63 (standard deviation 0.13). The scores varied from 0.29 to 0.86. The frequency with which the abstracts met the criteria is shown in Table 2. Most of the abstracts contained some information from each of the eight basic categories of an abstract. Although the abstracts addressed each category in general terms, there were deficiencies in particular areas.

At least some information on the research design was given in 25 (76%) of the abstracts; however, only 11 (44%) included specific technical descriptors

(e.g., randomized controlled trial, case-control study or survey). Fifteen abstracts (45%) did not describe the study setting. Twenty-nine (88%) gave some information on the subjects; however, 23 (79%) of them did not provide specific technical terms (e.g., random sample, convenience sample or consecutive sample) for the selection procedure. The study variables had to be inferred from the abstract in 17 cases (52%). Details concerning the data sources were absent from 12 (36%) of the abstracts.

The response and refusal rates were not indicated in 24 (92%) of the 26 abstracts in which they were considered applicable. Although 32 (97%) of the abstracts gave some information on the study results 21 (66%) did not provide suitable numerical data (e.g., probability values or confidence intervals) to support the findings.

In the 29 abstracts with conclusions the conclusions were not consistent with the results presented in 19 (66%), the study limitations were not mentioned in 25 (86%), there was no indication of how and to whom the findings might be useful in 12 (41%), and no recommendations for future research were given in 27 (93%).

Discussion

A mean score of 0.63 indicates that more than one-third of the information expected to be found in the abstracts was not present. Much of the missing information would have assisted the reader in determining the potential relevance of the study to his or her clinical practice. The omission of technical descriptors of research design, for example, makes it difficult to determine whether the study was performed with sufficient rigour to justify reading the whole article. Most of the abstracts failed to provide supporting data for the findings or to mention possible study limitations in the conclusions. These findings are consistent with those of Gotzsche,⁷ Evans and Pollock⁸ and Pocock, Hughes and Lee.⁹ Improvement is required in this area.

We were able to minimize the potential for error due to interrater variance by developing written evaluative criteria, conducting a pilot study of the developed checklist and having an independent third party serve as final arbiter of any disagreements. The fact that the interrater reliability in the actual study (0.82) was almost identical to that in the pilot study (0.81) suggested high interrater consistency.

Knowledge of information about the authors and the title of the article can influence the evaluation of an abstract. Blinding of the raters to these factors reduced the potential for bias.

The small sample of abstracts from 1 year of one journal was a limitation in making broad statements about the quality of abstracts. However, the risk of

bias due to temporal variations necessitated the sampling of only abstracts from articles published in 1 year. The quality of abstracts should be assessed in other journals, for other years and among authors from different medical specialties.

A further limitation was that the contents of the abstract and the article were not compared to determine whether the quality of the abstract reflected the quality of the article. Studies should be done in this area.

The omission of important information in traditional abstracts observed in our study reinforces the need for a more systematic format, such as that proposed for structured abstracts. In 1987 the Ad Hoc Working Group for Critical Appraisal of the Medical Literature¹⁰ proposed a structured abstract for clinical articles. It was felt that such a format would address perceived deficiencies in the quality of traditional abstracts. The following year Mulrow, Thacker and Pugh¹² developed similar guidelines for review articles. Since then a number of medical journals have given authors the option of submitting structured abstracts and have invited critical comments from their readers.^{4,5,15} Haynes and collaborators¹⁶ reviewed the acceptability and feasibility of structured abstracts and described revised guidelines.

Since the structured format was first proposed a number of journals, including the *Journal of General and Internal Medicine*,¹ *Chest*⁴ and the *British Medical Journal*,¹⁵ have required its use. Reader response has been favourable, and editors have not voiced any complaints. Structured abstracts have been accepted at scientific meetings and are now included in MEDLINE. Modifications in reporting statistics in structured abstracts have followed the revised guidelines proposed by the International Committee of Medical Journal Editors.¹³ As of January 1991 *CMAJ* is using the structured format proposed by Haynes and collaborators¹⁶ for abstracts of original research articles and review articles.¹⁷

Structured abstracts should not be considered a cure-all. Haynes and collaborators¹⁶ cited examples in which the structured abstract contained more rigorous design features than the article,^{18,19} omitted key details about loss to follow-up in the results²⁰ and provided conclusions that were not supported by the data presented.¹⁹ "Investigators with weaker studies tend to overstate the strength of their designs and conclusions, either by omission of important weaknesses or commission of inaccuracies in describing the features of their investigation."¹⁶

Although structured abstracts do not ensure accuracy, reporting errors may be more obvious and fewer than in traditional ones. This hypothesis, however, has not been examined. We recommend that studies be undertaken to determine whether the introduction of structured abstracts has improved abstract quality.

References

1. More informative abstracts [E]. *J Gen Intern Med* 1989; 4: 76
2. Fletcher RH: Writing an abstract [E]. *J Gen Intern Med* 1988; 3: 607-609
3. Pitkin RM: The importance of the abstract [E]. *Obstet Gynecol* 1987; 70: 267
4. Soffer A: Abstracts of clinical investigations: a new and standardized format [E]. *Chest* 1987; 92: 389-390
5. Squires BP: Structured abstracts. *Can Med Assoc J* 1990; 142: 703
6. Haynes RB, McKibbon KA, Walker CJ et al: Online access to MEDLINE in clinical settings: a study of use and usefulness. *Ann Intern Med* 1990; 112: 78-84
7. Gotzsche PC: Methodology and overt and hidden bias in reports of 196 double-blind trials of nonsteroidal antiinflammatory drugs in rheumatoid arthritis. *Controlled Clin Trials* 1989; 10: 31-56
8. Evans M, Pollock AV: Trials on trial: a review of trials of antibiotic prophylaxis. *Arch Surg* 1984; 119: 109-113
9. Pocock SJ, Hughes MD, Lee RJ: Statistical problems in the reporting of clinical trials: a survey of three medical journals. *N Engl J Med* 1987; 317: 426-432
10. Ad Hoc Working Group for Critical Appraisal of the Medical Literature: A proposal for more informative abstracts of clinical articles. *Ann Intern Med* 1987; 106: 598-604
11. Sacks HS, Berrier J, Reitman D et al: Meta-analyses of randomized controlled trials. *N Engl J Med* 1987; 316: 450-455
12. Mulrow CD, Thacker SB, Pugh JA: A proposal for more informative abstracts of review articles. *Ann Intern Med* 1988; 108: 613-615
13. International Committee of Medical Journal Editors: Uniform requirements for manuscripts submitted to biomedical journals. *Can Med Assoc J* 1988; 138: 321-328
14. Rosenthal R: *Meta-analytic Procedures for Social Research*, Sage, Beverly Hills, Calif, 1984: 55-62
15. Lock S: Structured abstracts [E]. *Br Med J* 1988; 297: 156
16. Haynes RB, Mulrow CD, Huth HJ et al: More informative abstracts revisited. *Ann Intern Med* 1990; 113: 69-76
17. Squires BP: Structured abstracts of original research and review articles [E]. *Can Med Assoc J* 1990; 143: 619-622
18. Nordstrom DM, West SG, Anderson PA et al: Pulse methotrexate therapy in rheumatoid arthritis. A controlled prospective roentgenographic study. *Ann Intern Med* 1987; 107: 797-801
19. Parving HH, Hommel E, Smidt UM: Protection of kidney function and decrease in albuminuria by captopril in insulin dependent diabetics with nephropathy. *Br Med J* 1988; 297: 1086-1091
20. Cummings SR, Coates TJ, Richard RJ et al: Training physicians in counselling about smoking cessation. A randomized trial of the "Quit for Life" program. *Ann Intern Med* 1989; 110: 640-647